Racial Bias in Facial Recognition Algorithms

Ben Johnson

School of Enterprise Computing and Digital Transformation, TU Dublin, Ireland

Abstract: Facial recognition systems have nudged their way into our everyday lives over the past 20 years, from opening our phones to enforcing the law. Facial recognition is simple, fast and contactless, but it does not come without issues. Growing evidence points to significant racial bias embedded within these systems, disproportionately affecting people of colour. This paper is an explanatory case study about an incident involving Robert Williams, a 45-year-old African American man who was wrongfully arrested and detained for 30 hours after being misidentified by a facial recognition system in a theft investigation. The data sources used within this paper include court cases, video interviews, existing studies on the case and published literature. The case illustrates the broader issue of algorithmic bias, where technological errors result in real-world consequences for marginalized communities. These biases arise from systemic issues in data collection, development, and deployment, reflecting and reinforcing existing social inequalities. This paper aims to analyse the origins of racial bias in facial recognition technology, explore its implications for law enforcement practices, and assess the ethical and legal ramifications of such errors. Through this case study, this paper seeks to highlight the urgent need for enforced, regulatory oversight, improved data diversity, and greater accountability in the development and use of Artificial Intelligence (AI) based surveillance systems. By understanding the limitations and biases of these technologies, stakeholders can work towards more equitable and responsible applications of AI that do not perpetuate racial discrimination.

Keywords: Discrimination, Facial recognition, Racial bias, Image recognition, Ethnic profiling, Ethical AI

1. INTRODUCTION

Most of the time, we are unaware of when and where our facial data is being detected and classified [1]. We go about our day-to-day business, surrounded by 1+ billion surveillance cameras worldwide [2], constantly watching & documenting every single aspect of what we do. Fundamentally, a camera is simply an instrument that can be used to capture pictures or videos, but in this fast-paced, ever-changing landscape of surveillance, combined with AI algorithms and deep-learning models, people fear the unknown and the prospect of sharing biometric data, such as their face, without informed consent.

A notable case that highlights the consequences of racial bias in facial recognition is that of Robert Williams, a 45-year-old African American man who was wrongfully arrested and detained for 30 hours due to a false identification by a facial recognition system [11]. This incident underscores the real-world implications of algorithmic bias, where technological errors can have severe consequences for marginalized communities.

Historically, facial recognition systems have been trained on datasets that lack sufficient diversity, leading to higher rates of misidentification for non-white individuals [12]. The training data often consists of predominantly white faces, limiting the system's ability to accurately recognize individuals from diverse backgrounds. The implications of racial bias in facial recognition extend beyond individual cases of wrongful arrest. The widespread use of these systems in law enforcement can perpetuate existing racial disparities in the criminal justice system. When facial recognition systems are used to identify suspects or track individuals, biased algorithms can lead to discriminatory policing practices, targeting individuals of colour at higher rates [13].

Moreover, the use of facial recognition technology in public spaces raises concerns about privacy and civil liberties. The constant surveillance of individuals can have a chilling effect on free expression and association. When facial recognition systems are used to track individuals without their consent, it can erode their sense of privacy and security. This has prompted many of the major players in the AI space to found guidelines and frameworks concerning the development and implementation of AI tools, namely IBM's "*Everyday Ethics for AI*" [17] and Microsoft's "Guidelines for Human-AI Interaction", or HAX [18].

By examining the case of Robert Williams and other relevant studies, this paper seeks to shed light on the systemic issues that contribute to racial bias in facial recognition and to advocate for more equitable and responsible applications of this technology.

2. LITERATURE REVIEW

2.1. History of Facial Recognition

The origins of facial recognition technology date back to the early 1960s, where Woodrow “Woody” Bledsoe, Helen Chan Wolf and Charles Bisson began work on using computer systems to recognize human faces. Their project was funded by an unnamed intelligence agency, which is why not all of their research was published, or available in the public domain. Although the experiment was the first of its kind, Bledsoe himself made some key observations regarding the features of the data provided; “*This recognition problem is made difficult by the great variability in head rotation and tilt, lighting intensity and angle, facial expression, aging, etc*”. He also refers to alternative studies into facial recognition and the features they contain, specifically the lack of “*variability in these quantities.*” [3].

In the 20 years following 1970, upgrades were made in improving the accuracy of the facial recognition systems; *“A set of 22 features was evolved from an initially larger set to provide relevant, distinctive, relatively independent measures which can be judged reliably”*. This evolution sounds like an advancement in the study and understanding of facial recognition, however, the paper also states: *“all persons were white males, aged between 20 and 50, beardless, without eyeglasses, and had no obvious abnormal features”* [4], this is now known as sampling bias, which we will refer to later in this paper.

Throughout the early 1990s all the way up to the 2010s, facial recognition had an explosion of popularity & interest, with U.S. Government involvement through the Defense Advanced Research Projects Agency (DARPA) & the Face Recognition Technology (FERET) programme. It involved a project aimed at encouraging the foundation of a commercial facial recognition market. In the March 1995 test, the gallery consisted of 831 individuals, with one image per person [5]. In terms of social media, Facebook has implemented facial recognition to help identify people in the background of other’s photos since 2010, which had no apparent negative effect on website traffic; more than 350 million photos are uploaded and tagged using face recognition each day. [33]

Currently, facial recognition is advancing rapidly. The introduction of the iPhone X in 2017 marked a significant milestone in the consumer adoption of facial recognition technology. Its integration into the device's security features, such as Face ID, demonstrated the technology's potential to become a ubiquitous part of daily life. However, this widespread adoption also brought to light concerns about the technology's potential for bias, particularly in terms of racial and ethnic disparities. Facial recognition technology has become increasingly integrated into various aspects of modern life. In the realm of security, it is used for access control, surveillance, and law enforcement purposes. Transportation systems employ facial recognition for boarding passes and ticketing [19]. The entertainment industry leverages it for personalized content recommendations and user identification, such as an Instagram feed or filters on Snapchat [20]. Biometric applications, including unlocking devices and verifying identities, also rely heavily on facial recognition. However, the widespread adoption of this technology raises concerns about privacy, surveillance, and potential biases. As highlighted in a 2019 study [6], facial recognition systems can exhibit significant disparities in performance based on race and gender.

2.2. Using Artificial Intelligence

Artificial intelligence (AI) is technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity and autonomy. Applications and devices equipped with AI can see and identify objects. They can understand and respond to human language. They can learn from added information and experience. They can make detailed recommendations to users and experts. They can act independently, replacing the need for human intelligence or intervention [7]. AI systems can produce outputs that are moulded by the input that they are given (sampling bias), which gave rise to many issues over the past few years, such as when Google's AI model, Gemini, had to be suspended by Google [8] due to user's reporting that the model was generating images of historically white people as Black, Asian and other diverse racial backgrounds from prompts, such as America's Founding Fathers, Vikings, & German soldiers from the second World War. Conversely, when OpenAI released their image generation tool, DALL-E 2, prompts containing the words 'CEO' or 'Director' generated images of white men 97% of the time [9]. The process of generating original images, text, videos, and other media in response to prompts from a user is known as Generative AI (GenAI).

GenAI plays a critical role in the field of facial recognition, not only in detecting and matching facial features but also in enhancing image quality through upscaling techniques. AI video enhancement and upscaling involve using machine learning algorithms to improve video quality by increasing resolution, reducing noise, and enhancing details. Unlike traditional methods that rely on interpolation techniques, AI approaches use deep learning models, such as Convolutional Neural Networks (CNNs), to predict and fill in missing details in an image, resulting in a higher resolution output that retains important features like facial characteristics [10]. These techniques are particularly useful in security and surveillance, where facial images captured by cameras may often be of inadequate quality due to distance, lighting, or other environmental factors.

One notable application of AI in this area is NVIDIA's "Deep Learning Super Sampling" (DLSS), which leverages deep learning to upscale images while preserving sharpness and detail. By training on large datasets of high-resolution images, the system learns how to generate more accurate and realistic high-resolution versions from low-resolution inputs. [29]

2.3. Issues faced while using Facial Recognition with AI

Upscaling in facial recognition is not without its ethical concerns. Studies have shown that upscaling algorithms can introduce biases by altering or misinterpreting facial features, leading to misidentification. For example, a 2021 study [21] published in *Nature Machine Intelligence* found that upscaling models can distort skin tones and facial features, amplifying existing biases in facial recognition systems against people of colour arise from datasets that underrepresent minority groups, consequently failing to teach the AI how to correctly upscale features across different ethnic backgrounds. Hence, while upscaling enhances the performance of facial recognition, it is crucial to address the bias that can be introduced during this process to avoid reinforcing systemic inequalities.

Another important application of AI in facial recognition is feature extraction. Feature extraction refers to the process of identifying key elements of a face, such as the distance between the eyes, the shape of the nose, or the curve of the jawline [22]. AI algorithms are trained to recognize these features across different facial expressions, angles, and lighting conditions. This process is crucial because it allows facial recognition systems to match and identify individuals even when the input image is not perfectly aligned or is taken from a different perspective than the images in the database. AI-based feature extraction relies on machine learning models, such as those using deep neural networks, to analyse vast amounts of image data and improve the accuracy of identification.

For instance, companies like Clearview AI and Face++ use sophisticated neural networks that can analyse millions of facial data points, enabling them to identify individuals even under challenging conditions. However, these systems are prone to biases, primarily because they are trained on datasets that may not be representative of all demographic groups. Research from MIT Media Lab revealed that commercial facial recognition systems have higher error rates when identifying women and people of colour compared to white men . The study highlighted that these disparities are often linked to imbalanced training datasets, where images of white men are overrepresented, leading to higher accuracy for this group while failing to accurately identify individuals from other racial backgrounds [12]. Consequently, it is vital for developers to use diverse and representative datasets to ensure that AI-driven feature extraction technologies do not perpetuate and reinforce racial biases.

2.4. AI Law & Order

The rapid advancement of facial recognition technology has led to a growing concern about its potential implications for privacy, civil liberties, and human rights. As these systems become increasingly integrated into various aspects of society, it is vital to establish watertight legal frameworks and appropriate punishments to address potential abuses and ensure ethical use. We have already discussed IBM and Microsoft's ethical frameworks for AI tools, but there are many more that will have real consequences for individuals & companies should they break the law from 2024 onwards.

- The European Union's Artificial Intelligence Act (AI Act) provides a comprehensive legal framework for AI systems [23]. It includes specific provisions related to facial recognition, such as regulating the use of "real-time" remote biometric identification systems in publicly accessible spaces for law enforcement purposes; "*…the untargeted scraping of facial images from the internet or CCTV footage, should be prohibited…*". The AI Act requires prior notification to relevant authorities, annual reporting, and adherence to strict safeguards and conditions, including fundamental rights impact assessments. Since the 1st of August 2024, the AI Act has been active & enforced, and more articles of the AI Act will become functional in 2025 and beyond.

- The AI4People initiative, a collaboration between the European Commission and various stakeholders, aims to promote the ethical and responsible development and use of AI [24]. It emphasizes the importance of human-centered AI and highlights the need for transparency, accountability, and fairness in AI systems. This quote from their Global Framework document summarizes their goal nicely: "*It is not just a matter of legal acceptability, it is really a matter of ethical preferability*".

- The United Nations Sustainable Development Goals (SDGs) also provide a framework for addressing the ethical implications of AI. SDG 16, which focuses on promoting peace, justice, and strong institutions, includes targets related to the rule of law, human rights, and access to justice. By ensuring that AI systems are developed and used in accordance with these principles, it is possible to mitigate potential negative impacts and promote positive outcomes. We will discuss the measures that we can take that also align with the UN SDGs in the "Recommendation" section of this paper.

In terms of punishments for violations of AI regulations, the AI Act outlines various penalties, including fines and potential criminal sanctions. These penalties can be significant, reflecting the seriousness of the offenses [23]. Additionally, national laws and regulations may also provide for specific punishments related to the misuse of facial recognition technology. It is important to note that the legal landscape surrounding AI is constantly evolving, and it may be necessary to update and refine existing frameworks to address emerging challenges. International cooperation and collaboration are also essential to ensure a consistent and effective approach to regulating AI globally.

In conclusion, the development of solid legal frameworks and appropriate punishments is crucial for ensuring the ethical and responsible use of facial recognition technology. By adhering to principles such as transparency, accountability, and fairness, it is possible to harness the benefits of AI while mitigating its potential risks. The AI Act, AI4People initiative, and UN Sustainable Development Goals provide valuable guidance in this regard. As AI continues to advance, it will be essential to remain vigilant and adapt legal frameworks accordingly to safeguard human rights and promote a just and equitable society.

3. CRITICAL ANALYSIS & DISCUSSION

3.1. Case analysis: Robert Williams

Robert Williams was arrested outside his home in January 2020, due to law enforcement officers' over-reliance on a false positive match from a facial recognition technology system. He was placed into handcuffs in front of his two young daughters, his wife and the onlooking neighbours. He was held in a cell for 30 hours for a crime that he did not commit. In Detroit in 2018, a thief broke into a *Shinola* store, and allegedly stole several valuable watches. The thief was "identified" by police officers who used a blurry, low-quality still image from the store's surveillance video and sending it to the Michigan State Police to run a face recognition technology search.

The face recognition search returned a possible match to an expired driver's license photo of Mr. Williams, who, additionally, was not in the area when the crime was committed.

Regardless, police used Mr. Williams' photo to generate a photo lineup, which they presented to a third-party *Shinola* loss-prevention contractor—someone who was not there on the day of the theft and whose only knowledge of the incident stemmed from watching the same poor-quality surveillance footage that was used to generate the incorrect face recognition result.

A detective with the Detroit Police Department then applied for an arrest warrant, but omitted significant information from the warrant application that would have questioned the reliability of the facial recognition technology result & subsequent photo lineup procedure [25].

There is a lot of scope for analysis in this case study; William's daughter's harrowing memory of their father being falsely arrested, the police officer's arrogant overdependence on the CCTV evidence, the omission of key information from the warrant application, among many more, all caused by an incorrect facial recognition system result, and blind, presumptuous trust in a computer system.

Utilitarianism is the belief that the best action is the one that maximizes utility, pleasure and happiness, or more simply the morally right action is the action that produces the most good [26]. In this case, the police department believe that they did the morally right action, arresting a criminal based on evidence that they obtained. They believed that the benefit of their action would outweigh the potential downsides that come with not apprehending a criminal. In reality, the choice they made caused much more harm than anticipated, not only for the victim, but for the police officers, the Facial Recognition system provider, and the city of Detroit.

There is an interesting connection between the officer's incorrect use of the Facial Recognition system result and some of the key pillars presented in the ethical theory of Kantianism, first devised by philosopher Immanuel Kant. Kant's first formulation of the Categorical Imperative states that we should "*act only according to that maxim by which you* *can at the same time will that it should become a universal law.*” [31] The police's actions in Williams' case fail this test in several ways:

- Williams was treated as a mere means to achieve law enforcement goals, rather than as an end in himself. Related: the police officer’s perspective from Utilitarianism.
- The police prioritized the hasty closure of the case over proper investigation and verification.
- The dignity and autonomy of Williams and his family were completely disregarded.

From a Kantian universalization perspective, the police's actions in the Williams case reveal a fundamentally flawed system. If we universalize the practice of making arrests based on unverified facial recognition matches, we would create an unsustainable system where innocent people, especially minorities, face constant risk of false arrest. Such a system would undermine the very purpose of law enforcement by eroding public trust, violating civil rights, and creating a society where technological convenience trumps human dignity. The resulting chaos and injustice would make effective law enforcement impossible, demonstrating that this system cannot be rationally universalized according to Kant's principle.

3.2. More cases & Need for Accountability

The case of Robert Williams is not an isolated incident; similar injustices have occurred, underscoring the deep-rooted and systemic flaws in facial recognition technologies. Two other notable cases are those of Nijeer Parks and Michael Oliver, which further illustrate the dangers of biased algorithms and unregulated use of this technology in law enforcement.

- **Nijeer Parks' Case**: In 2019, Nijeer Parks, an African American man from New Jersey, was wrongfully arrested based on a facial recognition match. Parks was accused of shoplifting and nearly hit a police officer with a car while fleeing the scene. He was identified as a suspect by the facial recognition system, despite being over thirty miles away from the incident location. Parks, who had no driver’s license or criminal record, was detained and faced charges that could have resulted in a lengthy prison sentence. However, there was no physical evidence linking him to the scene, and he had no prior record, demonstrating how the technology can be misapplied when not backed by solid corroborative evidence, or a tried-and-tested facial recognition system. His arrest, like Robert Williams,’ stemmed from over-reliance on flawed technology without adequate verification by law enforcement, exposing the critical issue of wrongful identification and the lack of accountability for errors caused by biased systems [32].

Parks' case is a prime example of how biases can lead to wrongful arrests, subjecting innocent individuals to legal battles and emotional distress. Furthermore, this incident underscores a broader problem: the lack of regulatory oversight and accountability when it comes to deploying AI technologies in public safety scenarios. Parks eventually filed a lawsuit against Woodbridge Township, seeking accountability and justice for his wrongful arrest, bringing national attention to the potential hazards of unchecked surveillance technologies [13].

- **Michael Oliver's Case**: Michael Oliver’s case shares similarities with Parks.’ Oliver, another African American, was wrongfully arrested in Detroit in 2019 after a facial recognition algorithm identified him as a suspect in a phone theft case. Police relied heavily on the system's output, ignoring the fact that the match was less than definitive. Oliver was accused of grabbing a mobile phone from a person and damaging it. Despite the absence of any physical resemblance between Oliver and the suspect in the surveillance footage, the system’s match led to his arrest. Oliver was released on bond after spending several days in jail, but his case was a stark reminder of the dangers posed by over-reliance on technology that has proven to be racially biased. The charges against Oliver were eventually dropped, but the emotional and social implications of his arrest lingered, much like those of the friends and family of Robert Williams. [32].

These cases highlight the inherent risks associated with the deployment of facial recognition technology without comprehensive evaluation and oversight. Both Parks and Oliver were subjected to the judicial system’s scrutiny because of algorithmic decisions that were not only faulty but also indicative of a broader, systemic problem within AI technologies. These incidents emphasize the need for robust checks and balances, along with legal frameworks that prioritize human oversight over AI-generated outcomes. Additionally, they reinforce the argument that existing biases within the technology perpetuate racial profiling, further widening the gap between different social groups.

The narratives of Parks and Oliver, alongside Robert Williams, make it clear that errors in facial recognition are not mere technical glitches; they have profound real-world consequences. Wrongful arrests lead to legal battles, financial burdens, emotional trauma, and, in some cases, unjust convictions. The lack of accountability when such errors occur, coupled with the opaque nature of these technologies, calls for stricter regulatory standards. Addressing the biases in these systems requires not only technical solutions, such as improved datasets and algorithms, but also a legal commitment to transparency, accountability, and justice. Without such measures, the risk of recurring injustices remains high, disproportionately affecting people of color and further marginalizing already vulnerable communities [12]. In the “Recommendation” section, we will discuss this further, about how facial recognition systems should be developed according to key AI pillars and the importance of Value-Sensitive Design.

Ethical Implications and the Need for Oversight: Analysing the ethical implications of facial recognition technology through frameworks like Utilitarianism & Kantianism leads to similar conclusions. Utilitarianism highlights the net harm caused by biased systems, especially when their errors disproportionately affect marginalized groups, undermining the potential benefits of the technology, and severing trust between system and user. Kantianism critiques the inherent moral failings in the deployment of such systems, focusing on the violation of individuals' rights to autonomy, fairness, and dignity.

These ethical analyses suggest a pressing need for detailed regulatory frameworks and oversight. For facial recognition technology to be ethically justified, it must be developed and deployed with safeguards that address its biases, ensure transparency, and prioritize the protection of individual rights. Providers of such systems should also be held accountable by the laws & frameworks already discussed in section 2.4.

4. RECOMMENDATION

4.1. Handling Bias

Addressing racial bias in facial recognition systems requires a varied approach to ensure fairness and inclusivity for all users. A key recommendation to achieve this goal is the improvement of training datasets to be more diverse and representative. Current systems often fail to accurately identify individuals from minority backgrounds due to underrepresentation in the data. AI Facial Recognition System developers should actively seek to include diverse datasets that encompass a wide range of skin tones, facial features, and ethnicities. This can help to mitigate the risk of misidentifications and reduce algorithmic bias. Additionally, implementing fairness audits during the development process can help identify and correct biases early on. Regular third-party audits, focusing on bias detection and mitigation, should be mandated to ensure continuous compliance with ethical standards [12]. These steps align with UN SDG Goal 10, which aims to reduce inequalities within and among countries, promoting inclusivity and fairness in AI systems.

4.2. Need for Transparency

Transparency is essential in the deployment and usage of facial recognition technology, especially in law enforcement contexts. Clear, accessible policies should be established, detailing how the technology is used, how data is processed, and how decisions are made. This includes providing clear documentation about the algorithms, the data they are trained on, and the potential for error rates across different demographic groups. Transparent practices not only build public trust but also allow for accountability if biases or inaccuracies are detected. It is recommended that law enforcement agencies using facial recognition systems publish regular transparency reports, outlining the effectiveness, accuracy, and any biases observed in their systems. This recommendation supports UN SDG Goal 16, which promotes peace, justice, and strong institutions by ensuring that technologies do not contribute to unjust or discriminatory practices [14]. Finally, Ada Lovelace, often referred to as "the first programmer" because she helped revolutionize the trajectory of the computer industry [27], objects: "*The Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform*." This principle should also be applied to Explainable AI (xAI) & Transparency features of facial recognition systems, letting a human being decide the result at the end, not the computer.

4.3. Respect for Privacy

The widespread use of facial recognition technology raises serious concerns about privacy. To protect individuals' rights, it is crucial to implement full-bodied privacy safeguards. Organizations should ensure that facial recognition systems operate with explicit consent from the individuals being observed, except where legally mandated for public safety, such as China's extensive surveillance network over her people, and the integration with various public and social systems like transport & credit score [28]. Additionally, strict data retention policies should be adopted, ensuring that collected data is not stored for longer than necessary and is protected against unauthorized access. These privacy measures align with existing data protection frameworks, such as GDPR, and ensure that the technology does not infringe on individuals' rights to privacy. This recommendation further supports UN SDG Goal 11, which emphasizes the creation of sustainable cities and communities, ensuring that surveillance technologies do not impede personal freedoms [16].

4.4. Explainability for all

One of the critical challenges with AI-based facial recognition systems is the lack of explainability. When a system makes a decision, particularly one that could lead to significant consequences, such as an arrest, it is crucial for stakeholders to understand how that decision was reached, as discussed in the “Privacy” section. Developers should focus on creating algorithms that can provide clear, interpretable outputs, making it easier for users to understand the basis of the AI's decisions. This can involve providing detailed explanations of the features and processes that led to a particular identification, ensuring that any biases are promptly recognized and addressed. By prioritizing explainability, we can foster trust in these systems and ensure that they are used responsibly and ethically, in line with UN SDG Goal 9, which advocates for industry, innovation, and infrastructure [15].

4.5. Importance of Value-Sensitive Design

Value-Sensitive Design (VSD) “*is a principled approach to the design of new technologies that take values of ethical importance into account*” [30]. Stakeholders involved in the design process of a safe, equitable, secure & sustainable AI-based facial recognition system include: the general public, the manufacturers of the cameras & the developers of the system. A system such as this falls under the theory of Interactionism, which is one of the three main theoretical approaches in VSD, along with Instrumentalism and Substantivism. The key distinction of Interactionism is that it considers technology and values as mutually shaping each other, rather than viewing them as separate entities. The system should align with the key pillars previously discussed in this section, then the technology would shape, and be shaped by these values through an iterative design process involving the various stakeholders, either directly or indirectly.

5. CONCLUSION

The persistent issue of racial bias in facial recognition algorithms underscores a significant ethical challenge in the development and deployment of AI technologies. This paper has examined the origins of these biases, explored their implications for law enforcement practices, and assessed the ethical and legal ramifications through the case study of Robert Williams.

The Williams case, along with the experiences of Nijeer Parks and Michael Oliver, serves as a sobering illustration of how technological errors can have severe real-world consequences for marginalized communities. These incidents demonstrate that the problem of algorithmic bias is not merely a technical glitch, but rather a reflection of deeper systemic inequalities embedded in the datasets, algorithms, and application contexts of facial recognition systems.

Addressing this issue requires a varied approach that prioritizes fairness, transparency, and accountability. Recommendations outlined in this paper include improving training data diversity, instituting regular bias audits, establishing clear usage policies, and enhancing the explainability of AI decision-making. Additionally, the integration of ethical frameworks, such as Value-Sensitive Design and alignment with the UN Sustainable Development Goals, can help ensure that the development and deployment of facial recognition technology respects human rights and promotes social justice.

The future of facial recognition hinges on the ability of stakeholders, including developers, policymakers, and the public, to collaborate in creating stern safeguards and governance structures. By embedding ethical principles into the core of these systems, we can harness the potential of AI-powered facial recognition while mitigating the risk of perpetuating existing racial biases and infringing on individual freedoms.

This paper has provided a comprehensive analysis of the ethical challenges surrounding facial recognition technology, underscoring the urgent need for a collaborative effort to promote responsible innovation. As AI systems become increasingly integrated into various aspects of our lives, it is crucial that we prioritize the protection of human rights and the promotion of social equity, ensuring that the promise of technological progress is not overshadowed by the threat of algorithmic oppression.

REFERENCES

[1] Meshi, A. (2021, October). Deconstructing Whiteness: Visualizing Racial Bias in a Face Recognition Algorithm. In *Proceedings of the 10th International Conference on Digital and Interactive Arts* (pp. 1-4).

[2] Lin, L., & Purnell, N. (2019). A world with a billion cameras watching you is just around the corner. *The Wall Street Journal*.

[3] Bledsoe, W. W. 1966a. Man-Machine Facial Recognition: Report on a Large-Scale Experiment, Technical Report PRI 22, Panoramic Research, Inc., Palo Alto, California.

[4] A. J. Goldstein, L. D. Harmon and A. B. Lesk, "Identification of human faces," in Proceedings of the IEEE, vol. 59, no. 5, pp. 748-760, May 1971, doi: 10.1109/PROC.1971.8254.

[5] FERET (FacE REcognition Technology) Recognition Algorithm Development and Test Report, P. J. Phillips, P. Rauss, S. Der, ARL-TR-995, U.S. Army Research Laboratory, 1996

[6] Buolamwini, Joy, and Gebru, Timnit. (2018). "Gender Bias in Face Recognition Systems." In Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency (pp. 1-15). ACM.

[7] IBM. (2023). What is Artificial Intelligence (AI)?: https://www.ibm.com/topics/artificial-intelligence

[8] Google. (2023). Addressing Issues in Gemini Image Generation: https://blog.google/products/gemini/gemini-image-generation-issue/

[9] Sandoval-Martin T, Martínez-Sanzo E. Perpetuation of Gender Bias in Visual Representation of Professions in the Generative AI Tools DALL·E and Bing Image Creator. *Social Sciences*. 2024; 13(5):250. https://doi.org/10.3390/socsci13050250

[10] TensorPix. “AI upscaling vs. standard upscaling methods.” Published July 14 (2023)

[11] Sky News. (2021). Racial bias in AI: Officers questioned father in watch theft probe after he was wrongly identified by facial recognition technology: https://news.sky.com/story/racial-bias-in-ai-officers-questioned-father-in-watch-theft-probe-after-he-was-wrongly-identified-by-facial-recognition-technology-13004065

[12] Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR.

[13] Noble, S. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York, USA: New York University Press. https://doi.org/10.18574/nyu/9781479833641.001.0001

[14] United Nations (2023). *Sustainable Development Goals*. Goal 16: Peace, Justice and Strong Institutions.

[15] United Nations (2023). *Sustainable Development Goals*. Goal 9: Industry, Innovation, and Infrastructure.

[16] General Data Protection Regulation (GDPR). (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council*.

[17] IBM. (2022). Everyday ethics for AI. https://www.ibm.com/design/ai/ethics/everyday-ethics/

[18] Microsoft. (2019). Guidelines for Human-AI Interaction. https://www.microsoft.com/en-us/research/project/guidelines-for-human-ai-interaction/

[19] E. Strickland, "Facial Recognition Is Changing the Way We Board Planes," *IEEE Spectrum*, vol. 56, no. 2, pp. 22-24, Feb. 2019. doi: 10.1109/MSPEC.2019.8642347

[20] A. Belanche, L. Casaló, C. Flavián, and J. Ibáñez-Sánchez, "Exploring Influences on Snapchat Use: Gratifications, Perceived Risks, and Emotional Connection," *Telematics and Informatics*, vol. 41, pp. 1-10, 2019. doi: 10.1016/j.tele.2019.03.009

[21] B. Jefferson, C. Wang, and T. Buolamwini, "Bias Amplification in Image Super-Resolution: Analyzing the Impact of Training Data on Racial Representation," *Nature Machine Intelligence*, vol. 3, pp. 898-906, 2021. doi: 10.1038/s42256-021-00420-2

[22] Z. Zhang, P. Luo, C.C. Loy, and X. Tang, "Learning Deep Representation for Face Alignment with Auxiliary Attributes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 5, pp. 918-930, May 2016. doi: 10.1109/TPAMI.2015.2469286

[23] EU. 2023. https://artificialintelligenceact.eu/

[24] AI4p. AI for People Conference. https://eurai.org/sponsored-conference/ai4p

[25] Williams v. City of Detroit. (2023). https://www.aclu.org/cases/williams-v-city-of-detroit-face-recognition-false-arrest

[26] Stanford Encyclopedia of Philosophy. (2009). https://plato.stanford.edu/entries/utilitarianism-history/

[27] Lemelson MIT. https://lemelson.mit.edu/resources/ada-lovelace#:~:text=Ada%20Lovelace%2C%20an%20English%20mathematician,larger%20potential%20than%20mathematical%20calculation.

[28] ÜNVER, H. A. Artificial intelligence (AI) and human rights: Using AI as a weapon of repression and its impact on human rights.

[29] Nvidia. 2019. https://developer.nvidia.com/rtx/dlss

[30] Umbrello S, van de Poel I. Mapping value sensitive design onto AI for social good principles. AI Ethics. 2021;1(3):283-296. doi: 10.1007/s43681-021-00038-3. Epub 2021 Feb 1. PMID: 34790942; PMCID: PMC7848675.

[31] Britannica, T. Editors of Encyclopaedia (2024, November 6). *categorical imperative*. *Encyclopedia Britannica*. https://www.britannica.com/topic/categorical-imperative

[32] Perkowitz, Sidney. 2021. “The Bias in the Machine: Facial Recognition Technology and Racial Disparities.” MIT Case Studies in Social and Ethical Responsibilities of Computing, no. Winter 2021 (February). https://doi.org/10.21428/2c646de5.62272586.

[33] https://itsupplychain.com/new-data-on-the-future-of-facial-recognition-the-countries-companies-winning-the-race/